01 Sep 1993

Identification of Variegated Coloring in Skin Tumors: Neural Network vs. Rule-Based Induction Methods

Ajaya Durg

William V. Stoecker
Missouri University of Science and Technology, wvs@mst.edu

J. P. Cookson

Scott E. Umbaugh

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/chem_facwork/2412

Recommended Citation

A. Durg et al., "Identification of Variegated Coloring in Skin Tumors: Neural Network vs. Rule-Based Induction Methods," *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, no. 3, pp. 71-74, Institute of Electrical and Electronics Engineers (IEEE), Sep 1993.
The definitive version is available at https://doi.org/10.1109/51.232345

Identification of Variegated Coloring in Skin Tumors

Neural Network vs. Rule-Based Induction Methods

Ajaya Durg[1]
William V. Stoecker[2,3,4]
John P. Cookson[5]
Scott E. Umbaugh[6]
Randy H. Moss[1]
[1]Department of Electrical Engineering
University of Missouri-Rolla
[2]Department of Internal Medicine
University of Missouri-Columbia
[3]Department of Computer Science
University of Missouri-Rolla
[4]Stoecker & Associates
[5]National Library of Medicine
[6]Department of Electrical Engineering
Southern Illinois University

In 1985, *Ca—A Journal for Clinicians* publicized the "ABCD's" of malignant melanoma [1]. This article and a previous article in 1982 [2] had noted the sharp rise in malignant melanoma incidence, up over 15 times since the 1930s; the sharpest current increase for any cancer, except for lung cancer in women [1, 2, 3]. It was apparent from the photographs and attendant medical information that the classical concept of nodular malignant melanoma had become outdated. The classic deadly blue-black or brown growing nodule did not adequately describe early malignant melanoma. Now, variegated coloring is the term often used to describe the varied hues present in malignant melanoma, often including tan, brown, and red, and sometimes shades of pink and blue. Variegated coloring is believed to be one of the most predictive features for malignant melanoma, yet it remains undefined except by example [1, 2]. In this article, we describe the use of neural networks for automatic identification of variegated coloring.

Induction Methods

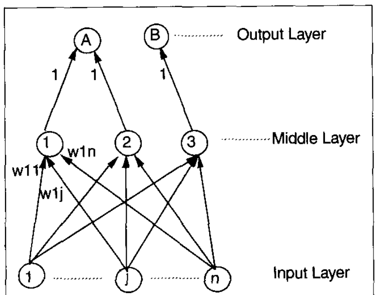
In a previous article [4], automatic induction, that is, the process of producing a general classification algorithm from a set of examples, was used to generate a classification rule to determine if variegated coloring was present in a given skin tumor image. The mechanism used in that research was an algorithm known as ID3, which operates by generating decision trees that are based on the input example data. ID3 can handle large amounts of data, with processing time growing only linearly with the complexity of the problem [5]. The algorithm, incorporated into the 1st Class Fusion software [6], was used in the previous article [4] to develop decision trees (coded in the C programming language) included in software used to classify skin tumors and skin tumor features. Neural networks are a mathematical model for the human brain. They may be considered to be another general category of induction methods. The basic neural network unit may be called the artificial neuron [7], with output firing or not firing depending on input, just as the human neuron does. Neural networks are based on the process of finding the best set of coefficients, or weight vectors, that minimize a given error function. Neural networks can thus be said to "learn," because they classify new input patterns into output patterns based on previous experience. In this way, neural networks perform induction.

Nestor Development System

For this research, the Nestor Development System (NDS) [8] was chosen for neural network implementation. NDS is a neural network development tool named after the mythical king Nestor, the prudent, eloquent, and wise king of Pylos. It is designed to solve pattern recognition problems of any kind. At the heart of NDS is a neural network, introduced in 1980, called a Restricted Coulomb Energy (RCE) network.

RCE Architecture

The RCE network consists of three processing layers; viz., an input layer, a middle or internal layer, and an output layer, as shown in Fig. 1. Cells (processing elements) in the input layer register the values of the input patterns. Each cell in the output layer corresponds to the category to which the input pattern belongs. Cells in the middle layer construct the mapping that ensures that the output cell for the correct category fires in response to a given input pattern. Each input cell is connected to all cells in the



ith cell in the middle layer is connected to the jth cell in the input layer via wij weight connection. Middle layer cells are connected to the output layer cells via a unit strength connection. Direction of the arrows indicate the flow of signals from input layer to output layer.

1. RCE network topology.

 0739-5175/93/$3.00©1993

middle layer by a weight connection, w_{ij}, connecting the *ith* middle-layer cell to the *jth* input layer cell. Each middle cell is connected to one and only one cell in the output layer via a unit-strength connection. Each middle-layer cell is committed to one category of patterns. All internal cells committed to a particular category of output pattern are connected to a single output cell. For example, suppose there are two output patterns, A and B, and three middle-layer cells, 1, 2, and 3, with cells 1 and 2 committed to pattern A, and cell 3 committed to pattern B. Middle layer cells 1 and 2 develop a connection to the output cell representing pattern A, and cell 3 develops a connection to the output cell representing pattern B , as shown in Fig. 1.

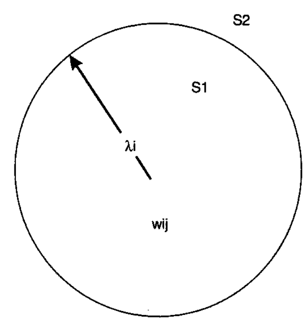
RCE is related to a model of neural information storage that likens the memory of the system to a collection of negative electrostatic charges at fixed sites in the feature space [8]. Clamping a pattern vector onto the input layer is analogous to introducing a positive test charge into a system of negative electrostatic charges. In the RCE network, the input vector will be finally stored in the form of weight connections associated with a particular middle-layer cell. This result is analogous to a positive charge moving under the influence of an electrostatic field until it comes to rest at one of the sites of negative charge. The RCE terminology reflects the fact that the memory sites are somewhat shielded or decoupled from one another [8]. The cell threshold associated with the middle-layer cells limits the attraction of any given memory to some local neighborhood surrounding the memory site. This decoupling allows training to affect only local areas of the memory.

Learning Scheme

The learning in the RCE network consists of two steps:

1. Committing the middle-layer cells.
2. Adjusting the threshold values.

First, input pattern values at each input cell are passed through the w_{ij} weight connections to the middle-layer cells, where w_{ij} is the weight connection from the *ith* middle-layer cell to the *jth* input layer cell. Outputs of middle-layer cells are passed through unit strength connections to the output layer cells. Each cell, i, in the middle layer has a transfer function of threshold value λ_i. The threshold function of the *ith* middle-layer cell can be visualized as a circular region of radius λ_i, with center, w_j, within the space of all possible signals input to the neural network, as shown in Fig. 2. (This is a simplification of the actual situation, as the actual transfer function covers a given complex region in n-space. Circular transfer functions would, of course, either overlap or leave uncovered regions, but we may distort the transfer function space to simplify the concept.) If a signal, s, entering a middle-layer cell, i, falls within the circular region S_i for the *ith* middle-layer cell, the *ith* middle-layer cell "fires." Then, the output cell connected to the *ith* middle-layer cell fires. Training modifies the connections of middle-layer cells to output cells, generally adding more middle-layer cells, and changes the threshold values or radii of middle-layer transfer functions as detailed below.



Transfer function of a middle layer cell could be represented as a disk-shaped region shown above. The cell fires for signals(S1) lying within the circular region, and does not fire for signals(S2) lying outside. S1, S2 are signals entering the middle layer cell from input layer.

2. Transfer function of a middle-layer cell.

There are three kinds of error signals, viz., +1, -1 and 0. If an output cell, representing a given category of patterns, is not firing (off) when it should be firing (on), an error signal +1 is generated for that output cell. If an output cell is firing (on) when it should be not firing (off), an error signal -1 travels from that cell back into the middle layer. If the output cell is firing/not firing properly, as it is required, then the error signal sent back to the middle layer is 0.

Error signal of +1:

An error signal of +1 traveling from the kth output cell into the system causes a new middle-layer cell to be committed. Its output is connected to the kth cell in the output layer, and its weight connections to the input layer assume the values of the current input pattern:

$$w_{ij} = f_i, \quad j = 1,..,n$$

Here, i is the new middle-layer cell being committed and $f_j(j = 1,..,n)$ is the current input pattern. The threshold value of the cell is set at some positive value, λ_i.

Error signal of -1:

If an error signal of -1 is sent from the kth output unit back into the system, then the system responds by reducing the threshold values (λ_i) of all the middle-layer cells that are active (firing) and connected to the kth output cell. This activity shrinks the circular transfer function region, shown in Fig. 2, and tends to turn the middle-layer cell off.

Error signal of 0:

In this case, the middle-layer cells connected to the output cell (which is sending 0 error signal) are not modified.

Comparison of RCE with Other Neural Network Models

The patterns that the system learns during training are stored in the weight vectors between the input layer and the output layer. The number of cells in the middle layer grows automatically as a function of the complexity of the problem. This automatic growth makes network development easier than for other network models in which a trial and error procedure is necessary to fix the number of hidden units or middle-layer cells.

An important property of the RCE network is that it can store an arbitrarily large number of patterns without degrading the performance of the system. The RCE network can store as many learned patterns as there are sites available in the space. In the RCE model, the number of middle-layer units grows with the input patterns so the memory, or sum of learned patterns, of the network is not limited by the architecture with which the initial training is started.

In the RCE networks, the cell threshold associated with the middle-layer cells limits the attraction of any given memory to some local neighborhood surrounding the memory site. This in turn allows training to affect only local areas of the memory. These local changes to memory can thus be made at arbitrary points in time, without the need to retrain the system on a global basis (i.e., repeat all instances of all categories from the original training set). This time-saving feature contrasts with other network models requiring the whole training process to be repeated when new patterns are introduced in the training set.

Nestor Learning System

The Nestor Learning System (NLS) is composed of multiple RCE networks along with a controller module. The controller integrates the responses of the component RCE networks into a system response and, on the basis of corrective feedback from an instructor, directs the training of the networks in the system. In the NLS, each RCE network processes a user-defined feature subset. By virtue of the feature subset it is processing, a particular RCE network may be trained to make unambiguous, correct decisions about patterns belonging to certain categories. For other categories, the partitioning may be such that no single RCE network has enough information to reliably classify the event. In such cases, an RCE network can develop category mappings that at least allow it to indicate likely pattern categories for the input. The NLS correlates the answers of several such networks in an attempt to identify the pattern. In addition, the controller can further partition feature subsets, constructing, in effect, new networks when no existing networks or combinations can make the required identification [8]. Thus, NLS combines several RCE networks into a complex neural network pattern recognition system.

Database for Detection of Variegated Coloring

Variegated coloring was determined for each of 250 tumors by two dermatologists as described in Ref. [4]. The two dermatologists agreed on presence or absence of variegation in each of these tumors. The dermatologists found 50 of these tumors to be variegated, and 200 to be non-variegated. This same data set was used in the current project, with the entire 250 member set divided into training and test sets by the same method as previously described. We summarize here the steps used in image processing to detect variegated coloring:

1. Color averaging
2. Feature masking
3. Color space segmentation
4. Object filtering
5. Object labeling
6. Feature identification

These steps are presented in detail in Ref. [9]. Features such as rulers and other misleading artifacts are masked out in the second step by filling this portion of the digitized image with zeros. Segmentation of the image into four colors by the color cell technique was used (as in Ref. [9]). The current research compares the use of neural networks to accomplish step 6, feature identification, to the heuristic and rule-based induction methods used in previous research.

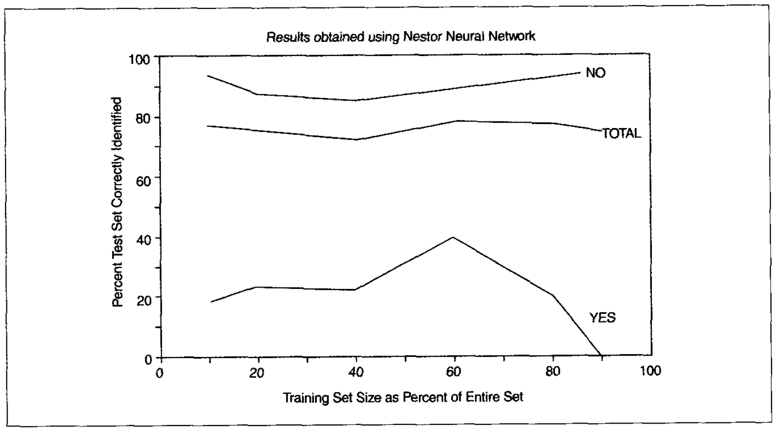
Data available to the neural network system included 18 inputs. Although discussed in Ref. [9], these inputs are reviewed here because of the importance in selection of data for the neural network. The first three inputs were the average tumor *relative* colors, expressed as the average tumor red, green, and blue pixel values, minus respectively, the average red, green, and blue background skin pixel values. Input number four was the ratio of feature-free area of the tumor to the entire area. Those features, including ulcer, crust, and similar features that were selected by the dermatologist, were not included in color processing for the tumor. For inputs 5-16, for each of four colors in the final segmentation, the numbers were the variance of angle A, variance of angle B, and the number of pixels for each color. Angle A and Angle B together denote a point in the Newton-Maxwell color triangle. Angle A denotes the angle between the line running from the left (red) corner of the triangle and the base, and Angle B denotes the angle between the line running from the right (green) corner of the triangle and the base. The final two inputs were the number of color objects greater than the minimum size and the number of different colors (with a maximum of four). The minimum color object size was chosen to be about 2 mm^2, believed by dermatologists to be the smallest meaningful color object. The variances in inputs 5 to 16 were scaled by a factor of 10^4 to achieve 3 significant figures without the use of scientific notation. The entire data base consisted of 18 numbers and one word, either "yes" or "no," indicating presence or absence of variegated coloring, for each of 250 tumors.

Training/Test Set Paradigm Applied to Variegated Coloring Database

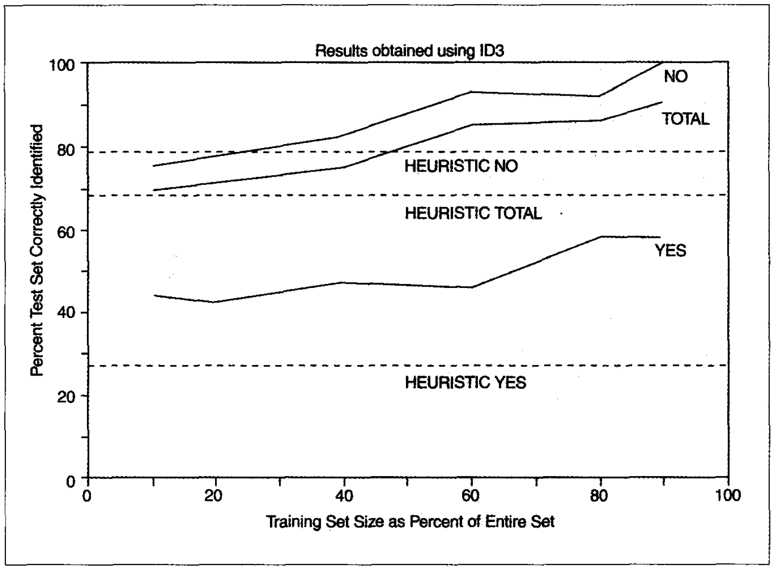
In developing the neural network experiment, each of the 250 tumor images was assigned to either the training or the test set. Using 10 percent of the entire set for training was accomplished by using every 10th tumor for training; using 20 percent of the entire set for training was accomplished by using every fifth tumor for training,... and using 90 percent of the entire set for training was accomplished by using every 10th tumor for testing. This scheme did not assure an identical percentage of variegated tumors in the training and test set for each run of the experiment, but it did assure that the experiment was run in the same way as for earlier experiments in testing 1st Class Fusion [9].

Results

Figure 3shows results obtained using the various percentages of the data for the training set. A peak diagnostic accuracy of 80 percent was obtained using the neural network. This peak accuracy is not very impressive considering the presence of 80 percent nonvariegated tumors. A default rule to obtain 80 percent accuracy would be simply to diagnose all tumors as nonvariegated. It is noted that moderate diagnostic accuracy is obtained with a small number of cases, but accuracy does not improve with a greater number of training cases. Diagnostic accuracy for the "yes" set containing tumors with variegated coloring increased up to a training-set size of 60 percent, and then decreased with a larger number of training cases. Figure 4 (from Ref. [9]) shows the training/test results with 1st Class Fusion, with lower diagnostic accuracy when a small percentage of the cases is used for training, but a higher accuracy when a high percentage of



3. Nestor results for automatic detection of variegated coloring with a data set of 250



4. ID3 results for automatic detection of variegated coloring with a data set of 250 tumors, varying training-set size as percent of entire set. The results for the expert-defined explicit heuristic are as shown, with 68 percent accuracy for the entire set. These results were obtained in the previous study, using the ID3 automatic induction algorithm [4].

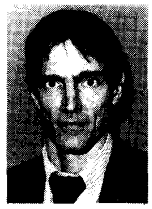
the cases is used for training. Figure 3 also shows a training/test accuracy of 68 percent, obtained with the expert-defined heuristics. This rate is a lower diagnostic accuracy than provided by either the neural network or the automatic induction models.

Discussion

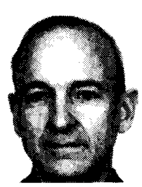
Recently, increasing attention has been devoted to medical application of implicit induction methods. These data-derived expert systems have rules that are not understood by experts, as opposed to explicit expert systems, with expert-defined heuristics. During the last decade, many man-hours were spent on explicit expert systems in various medical and non-medical fields, attempting to capture the expert's knowledge with rule-based systems. We have found our best expert-defined heuristics, both on the variegation problem and in skin cancer recognition in general, to be unsatisfactory for clinical application. In contrast, automatic induction methods, for example those obtained with the 1st Class Fusion ID3 algorithm, can provide high diagnostic accuracy with large training set sizes. Neural networks have been applied to other clinical problems [11-14], with at least one example of neural network clinical accuracy significantly exceeding that of physicians [14]. We noted limitations of performance for our neural network model. Improved performance might be obtained with a combination of several levels of Nestor neural networks or with preprocessing of the data. Results are presented here for only a single level of Nestor, without preprocessing, to provide direct comparison between a single level of Nestor neural network and an automatic induction rule-based algorithm. In our domain, for the diagnosis of variegated coloring, a single level of Nestor neural network method performs better than the ID3 rule-based automatic induction for small training-set sizes, and worse than rule-based automatic induction for larger training-set sizes. It is not possible to draw general conclusions from our comparison of automatic induction methods.

Ajaya Durg is currently pursuing the M.S. in electrical engineering at the University of Missouri-Rolla (UMR). He received the B.S. in electronics and communication engineering from Osmania University, India, in 1989. His areas of interest include computer architecture, VLSI design, computer networks, and neural network applications.

William V. Stoecker practices dermatology in Rolla, Missouri. He received the B.S. in mathematics from Caltech in 1968,



the M.S. in systems science from UCLA in 1970, and the M.D. from the University of Missouri-Columbia (UMC) in 1977. He is Clinical Assistant Professor of Internal Medicine-Dermatology at UMC and Adjunct Assistant Professor of Computer Science at UMR. He is Chairman of the American Academy of Dermatology Task Force on Computer Data Bases, which has developed diagnostic and therapeutic software for dermatologists. He is president of Stoecker & Associates, which develops medical computer vision systems. His interests include artificial intelligence in medicine, computer vision in medicine and diagnostic problems in dermatology. **Dr. Stoecker can be reached at 1100 W. 10th St., Rolla MO 65401; Tel.: (314) 364-0122; Fax: (314) 364-0129.**



John Cookson received the B.A. degree from Maryknoll College in 1961 and both bachelor's and master's degrees in electrical engineering from Catholic University in 1968 and 1971, respectively. He has also pursued additional technical studies at the University of Maryland. He was affiliated with the National Library of Medicine's Lister Hill Center for Biomedical Communications from 1980 to 1991 as a research engineer. Cookson is currently affiliated with the Library of Congress's Service for the Blind and Physically Handicapped, Washington, DC. His work centers on the exploration of new library access strategies for persons who are blind and physically handicapped.

Scott E. Umbaugh is currently Assistant Professor in Electrical Engineering at Southern Illinois University at Edwardsville. He received the B.S. degree from SIUE in 1982, the M.S. from the University of Missouri-Rolla in 1987, and the Ph.D. in electrical engineering from UMR in 1990. Umbaugh worked as a computer software/hardware engineer for ITT North Electric in Columbus, Ohio and McDonnell Douglas Corp. and Affinitec Corp. in St. Louis from 1981 through 1986. He is a member of several professional societies.

(continued on page 98)

(continued from page 74)

Randy H. Moss is currently Professor of Electrical Engineering at the University of Missouri-Rolla. He received the B.S. and M.S. degrees in electrical engineering from the University of Arkansas in 1975 and 1977, respectively. He received the Ph.D. degree in electrical engineering in 1981 from the University of Illinois. He joined the faculty of the University of Missouri-Rolla in 1981 as an assistant professor and was promoted to his current rank in 1991. He is a member of several professional societies, including IEEE.

References

1. **Friedman R, Rigel DS, Kopf AW:** Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *Ca—A Cancer Journal for Clinicians* 35(3):130-151, 1985.
2. **Kopf AW, Rigel DS, Friedman RJ:** The rising incidence and mortality rate of malignant melanoma. *J Dermatol Surg Oncol* 8:760-761, 1982.
3. **Silverberg E:** Cancer statistics, 1985. *CA* 35:19-35, 1985.
4. **Umbaugh SE, Moss RH, Stoecker WV:** Applying artificial intelligence to the identification of variegated coloring in skin tumors. *IEEE Engng Med Biol* 10(4):57-62, 1991.
5. **Quinlan JR:** Learning efficient classification procedures and their application to chess end games. In: Michalski RS, Carbonell J, Mitchell TM (Eds): *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufman, Los Altos, CA, 1983.
6. **Thomas W, Hapgood W:** *1st-Class Fusion*, 1st-Class Expert Systems, Inc., Wayland Mass., 1987.
7. **Devijver PA, Kittler J:** *Pattern Recognition: A Statistical Approach*, Prentice-Hall International, London, 1982.
8. *Neural Network Systems, An Introduction for Managers, Decision Makers and Strategists*, Nestor Inc., One Richmond Square, Providence, RI, 02906, 1988.
9. **Umbaugh SE, Moss RH, Stoecker WV:** Automatic color segmentation of images with application to detection of variegated coloring in skin tumors. *IEEE Engng Med Biol* 8(4):43-52, 1989.
10. **Stoecker WV:** Computer-aided diagnosis of dermatologic disorders. *Dermatol Clin N Amer* 4:607-625, 1986.
11. **Hudson DL, Cohen ME, Anderson MF:** Determination of testing efficacy in carcinoma of the lung using a neural network model. *Symp Comput Applic Med Care, 1988; Proc 12th Ann Symp*, Washington DC, 12:251-5, 1988.
12. **Kiraiwa A, Shimohara K, Tokunaga Y:** EEG topography recognition by neural networks. *IEEE Engng Med Biol* 9:39-42, 1990.
13. **Yoon YO, Brobst RW, Bergstresser PR, Peterson LL:** A desktop neural network for dermatology diagnosis. *Neural Networks* 1:43-52, Summer 1989.
14. **Baxt WG:** Use of an artificial neural network for the diagnosis of myocardial infarction. *Annals Internal Med* 115:843-848, 1991.